

PATENT

Atty Docket No.: 10008388-1

App. Ser. No.: 10/066,096

In the Claims

Please find below a listing of all of the pending claims. The statuses of the claims are set forth in parentheses.

1. (Currently Amended) A computerized method for measuring the a degree of coherence of the arrangement of nodes in a hierarchy comprising the steps of:

a) receiving a predetermined hierarchy of nodes arranged in a tree format with one or more subtrees, wherein a subtree of one of the nodes includes any descendant nodes in the hierarchy that stem from the one node, the hierarchy of nodes includes at least a first node and a second node, and the nodes in the hierarchy are associated with one another as one of a sibling node, a child node, and a parent node,

wherein the second node is a child node of the first node when the second node stems from the first node without any intervening node therebetween and the second node belongs in the subtree of the first node;

wherein the first node is a parent node of the second node when the second node is the child node of the first node and belongs in the subtree of the first node;
and

wherein the first node is a sibling node with the second node when the first and second nodes stem from a same parent node without any intervening node therebetween, and the first node and the second node belong to the subtree of the same parent node;

b) receiving a plurality of training cases that are filed under said the nodes in the hierarchy; and

PATENT

Atty Docket No.: 10008388-1

App. Ser. No.: 10/066,096

c) responsive thereto for determining a coherence measure of coherence, for at least one of the nodes in the hierarchy that has a local environment, by evaluating the training cases filed in the subtree of the at least one node with respect to the training cases filed in the local environment of the at least one node;

wherein the local environment of the at least one node includes any parent node of the at least one node, any nodes that are sibling nodes of the at least one node, and any nodes that belong to the subtrees under the sibling nodes of the at least one node.

2. (Currently Amended) The method of claim 1 wherein the step of determining a coherence measure of coherence includes the steps of

determining, for the subtree at of the at least one node, the number of the training cases filed in the subtree and the average prevalence of each feature in the training cases filed in the subtree;

determining, for the local environment of the at least one node, the number of the training cases filed in the local environment and the average prevalence of each feature in the training cases filed in the local environment;

determining predictive features that distinguish the subtree of the current at least one node from the local environment of the at least one node; and

generating a coherence value for the current at least one node based on the average prevalence of at least one of the predictive features.

PATENT

Atty Docket No.: 10008388-1

App. Ser. No.: 10/066,096

3. (Currently Amended) The method of claim 2 further comprising the steps of determining, for each said of the predictive features, the a degree of uniformity of the prevalence of the each predictive feature among the children subtrees of the at least one node; and wherein the step of generating a coherence value for the ~~current~~ at least one node is based on said the degree of uniformity and the average prevalence of the at least one predictive feature.

4. (Original) The method of claim 1 wherein the hierarchy of nodes includes a topic hierarchy; wherein the nodes are topics; and wherein the training cases includes one of labeled documents and feature vectors assigned to the topics.

5. (Previously Presented) The method of claim 2 wherein the predictive features include at least one of words, multi-word phrases, noun phrases, document length, file extension type, and other parameters related to documents.

6. (Currently Amended) The method of claim 2 5 wherein the step of determining the predictive features includes the step of computing at least one of information-gain metrics, mutual-information metrics, Chi-Squared, Fisher's Exact Test, lift, odds-ratio, word frequency among documents, and word frequency among all words in all of the documents.

7. (Current Amended) The method of claim 3 wherein the step of selecting features that are uniformly common includes

PATENT

Atty Docket No.: 10008388-1
App. Ser. No.: 10/066,096

the step of computing one of the metrics cosine-similarity, projection, and ChiSquared between the average feature prevalence vector and the vector of training case counts across the subtopics of the current at least one node, wherein the nodes in the hierarchy are topics and any child node of one of the nodes is also a subtopic of its parent node.

8. (Currently Amended) The method of claim 3 wherein the step of generating a ~~hierarchical coherence-number~~ the coherence value includes the step of generating a hierarchical coherence number by computing the average prevalence of the at least one predictive feature with the greatest degree of uniformity.

9. (Currently Amended) The method of claim 3 wherein the step of generating a ~~hierarchical-coherence-number~~ the coherence value includes the step of generating a hierarchical coherence number by computing a weighted-average of the average prevalence of at least two of the predictive features that are selected as both predictive and uniform.

10. (Currently Amended) The method of claim 9 ⁵ wherein the step of generating a ~~hierarchical-coherence-number~~ the coherence value includes the step of generating a hierarchical coherence number by computing a weighted-average of the average prevalence of the top k most prevalent of the predictive features that are selected as both predictive and uniform, wherein k is a predetermined positive integer.

PATENT

Atty Docket No.: 10008388-1
App. Ser. No.: 10/066,096

11. (Currently amended) The method of claim 9 10 wherein the weighted-average employs as the weighting schedule one of the negative exponential function $\exp(-l)$ and the inverse rank function $(1/l)$, where l is the ordered rank of the top k most prevalent of the predictive features that are selected as both predictive and uniform.

12. (Currently Amended) The method of claim 3 wherein the step of generating a ~~hierarchical-coherence-number~~ the coherence value includes the step of generating a hierarchical coherence number by computing the an average value of the average prevalence of the top k most prevalent of the predictive features that are selected as both predictive and uniform, wherein k is a predetermined positive integer.

13. (Currently Amended) The method of claim 2 wherein the step of generating a ~~hierarchical-coherence-number~~ the coherence value includes the step of generating a hierarchical coherence number by employing a maximum, over all of the predictive features, of a projection between the average feature prevalence vector and the vector of training case counts across ~~the~~ subtopics of the current at least one node, wherein the nodes in the hierarchy are topics and any child node of one of the nodes is also a subtopic of its parent node.

14. (Currently Amended) The method of claim 2 wherein the step of generating a ~~hierarchical-coherence-number~~ the coherence value includes the step of generating a hierarchical coherence number by employing a maximum average prevalence of the predictive features.

PATENT

Atty Docket No.: 10008388-1

App. Ser. No.: 10/066,096

15. (Currently Amended) The method of claim 1 further comprising the step of:
assigning an aggregate-coherence value to a node in the hierarchy, based on an aggregation function of said determined ~~measure of coherence~~ value over said ~~the~~ node and of descendants of said the node.
16. (Original) The method of claim 15 wherein the aggregation function includes one of a sum, average, weighted-average, minimum function, and maximum function.
17. (Currently Amended) The method of claim ~~[[1]]~~ 2 further comprising the step of:
using the coherence values of one or more nodes in the hierarchy to modify the structure of the hierarchy to improve the coherence of the hierarchy.
18. (Currently Amended) The method of claim ~~[[1]]~~ 2 further comprising the step of:
using the coherence values of one or more nodes to guide the selection of training cases for an automated classifier.
19. (Currently Amended) The method of claim ~~[[1]]~~ 2 further comprising the step of:
using the coherence values of one or more nodes to select a suitable classification technology to be employed to automatically classify items in the hierarchy.
20. (Currently Amended) A computerized apparatus for measuring the a degree of coherence of at least one considered node in a hierarchy of nodes that has associated therewith a subtree and a local environment in a the hierarchy comprising:

PATENT

Atty Docket No.: 10008388-1
App. Ser. No.: 10/066,096

a) a training case counter for determining the number of training cases under the subtree and the number of training cases for the local environment, the subtree includes any nodes in the hierarchy that stem from the at least one considered node, and the local environment includes any parent node from which the at least one node is stemmed directly and is thus a child node of the parent node, any sibling nodes that are stemmed directly from the parent node of the at least one node, and any nodes that stem from the sibling nodes of the at least one node;

b) a predictive feature determination unit for determining a set of predictive features that distinguish training cases of the subtree from documents of the local environment;

c) an average prevalence determination unit for determining for at least one feature the average prevalence under the subtree and the average prevalence for the local environment; and

~~c) a predictive feature determination unit for determining the set of predictive features that distinguish training cases of the subtree from documents of the local environment; and~~

d) a coherence assignment unit for generating a coherence metric number for each the at least one considered node based on at least one predictive feature.

21. (Currently Amended) The apparatus of claim 20 further comprising:

a subtopic uniformity determination unit for determining the uniformity of the distribution of said the predictive features among the children subtopics of the at least one considered node, wherein the nodes in the hierarchy are topics and any child node of one of the nodes is also a subtopic of its parent node;

PATENT

Atty Docket No.: 10008388-1
App. Ser. No.: 10/066,096

wherein the coherence assignment unit generates a coherence metric number based on at least one predictive feature that is determined to be uniformly distributed among said the children subtopics.

22. (Currently Amended) A computerized system for measuring the a degree of coherence of nodes in a topic hierarchy comprising:

a) a coherence analyzer unit for receiving the topic hierarchy and a set of labeled training cases filed under each of the nodes in the topic hierarchy and responsive thereto for determining, for at least one current node under consideration from the nodes in the topic hierarchy, a measure of coherence measure in the topic hierarchy of the at least one current node under consideration by evaluating the training cases and at least one feature under the a local environment of the at least one current node and by evaluating the training cases and at least one feature under the a subtree of the at least one current node under consideration;

wherein the subtree of the at least one current node under consideration includes any of the nodes in the topic hierarchy that stem from the at least one current node under consideration; and

wherein the local environment of the at least one current node under consideration includes any of the nodes in the topic hierarchy that stem from a parent node from which the at least one current node under consideration is stemmed directly.

PATENT

Atty Docket No.: 10008388-1
App. Scr. No.: 10/066,096

23. (Currently Amended) The system of claim 22 further comprising:

b) a user interface presentation unit coupled to the coherence analyzer unit for displaying ~~a measure of the~~ coherence measure for one or more current nodes under consideration.

24. (Currently Amended) The system of claim 22 further comprising:

b) a feature extractor coupled to the coherence analyzer for receiving a set of labeled documents and at least one feature guideline and responsive thereto for generating the set of labeled feature vectors.

25. (Currently Amended) The system of claim 22 wherein the coherence analyzer unit further comprises:

a_1) a training case counter for determining the number of training cases under ~~each node~~ a subtree of each of the nodes;

a_2) an average prevalence determination unit for determining the average prevalence for at least one feature under each of the node subtrees; and

a_3) a predictive feature determination unit for determining predictive features under each of the node subtrees; and

a_4) a coherence assignment unit for generating coherence metric number based on at least one of the predictive features.

26. (Currently Amended) The system of claim 25 wherein the coherence analyzer unit further comprises:

PATENT

Atty Docket No.: 10008388-1
App. Scr. No.: 10/066,096

a_5) a subtopic uniformity determination unit for determining the degree of uniformity in the distribution of one or more said of the predictive features among the children of the at least on current node;

wherein the coherence assignment unit generates a coherence metric number based on at least one uniform of the predictive features that is deemed uniform based on the determined degree of uniformity of the at least one uniformity.

27. (Currently Amended) A computerized method for measuring the a degree of coherence for one or more current nodes in a hierarchy of nodes comprising the steps of:

a) receiving a the hierarchy and the training cases filed into said the hierarchy;

b) determining a list of predictive features that distinguish documents of a subtree of a first one of the nodes in the hierarchy the current node's sub-tree from these from documents in the current first node's local environment, wherein the first node's subtree includes any nodes in the hierarchy that stem from the first node, and the first node's local environment includes any parent node from which the first node is stemmed directly, any sibling nodes that are stemmed directly from the first node's parent node, and any nodes that stem from the sibling nodes of the first node;

c) assigning a coherence value to the current first node -based on the list of predictive features and based on one or more of their degree of predictiveness, their the degree of prevalence, and their degree of uniformity, wherein the degree of uniformity reflects how evenly distributed said predictive features are among the children subtrees of the children nodes in the hierarchy that are directly stemmed from the first current node based on the training cases under each of the child-subtree subtrees of the children nodes.